DOCUMENT RESUME

ED 386 465 TM 023 812

AUTHOR Enright, Mary K.; And Others
TITLE A Cognitive Analysis of Solutions for Verbal, Informal, and Formal-Deductive Reasoning Problems. GRE Board Professional Report No. 90-04P.
INSTITUTION Educational Testing Service, Princeton, NJ. Graduate Record Examination Board Program.
SPONS AGENCY Graduate Record Examinations Board, Princeton, N.J.
REPORT NO ETS-RR-95-6
PUB DATE Feb 95
NOTE 58p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS College Entrance Examinations; *College Students; Deduction; Evaluation Methods; Higher Education; *Problem Solving; Test Construction; *Test Items; *Thinking Skills; Verbal Ability
IDENTIFIERS Analytical Tests; Formal Deductive Problems; *Graduate Record Examinations; Justification (Logic)

ABSTRACT

A previous study of new item types for the analytical measure of the Graduate Record Examinations (GRE) General Test found that the new items had many factors labeled verbal reasoning, informal reasoning, formal-deductive reasoning, and quantitative reasoning. The present study examined how processing differed for these item types in the context of a problem-space framework. Protocols of 16 graduate and undergraduate students solving a small set of items aloud were collected and examined according to problem representation and problem solution. The representation of formal-deductive items involved the use of meaning-reduced tokens and spatial diagrams. The units involved in the representation of informal reasoning and verbal reasoning item types included meaningful propositions and meaning-emphasized paraphrases. The order of processes of evaluation and justification was found to differ for formal-deductive items and other item types. Item solutions also varied in terms of the kinds of justifications that were offered by the examinees for accepting or rejecting options. These results illustrate how the addition of some item types to the GRE analytical measure will expand the variety of reasoning skills assessed. Implications of these results for cognitive models of reasoning are also discussed. Six tables and one figure illustrate the analysis. (Contains 36 references.) (Author/SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 386 465

GRE®
RESEARCH

# A Cognitive Analysis of Solutions for Verbal, Informal, and Formal-Deductive Reasoning Problems

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Mary K. Enright
Carol B. Tucker
and
Irvin R. Katz

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

February 1995

GRE Board Professional Report No. 90-04P
ETS Research Report 95-06

BEST COPY AVAILABLE

Educational Testing Service, Princeton, New Jersey

# A Cognitive Analysis of Solutions for Verbal, Informal, and Formal-Deductive Reasoning Problems

Mary K. Enright
Carol B. Tucker
and
Irvin R. Katz

GRE Board Report No. 90-04P

February 1995

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, NJ 08541

**********

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board Reports do not necessarily represent official Graduate Record Examinations Board position or policy.

**********

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, GRADUATE RECORD EXAMINATIONS, and GRE are registered trademarks of Educational Testing Service.

Copyright © 1995 by Educational Testing Service. All rights reserved.

Abstract

A previous study of new item types for the analytical measure of the GRE General Test found that the items loaded on three of four separable factors that were labeled verbal reasoning, informal reasoning, formal-deductive reasoning, and quantitative reasoning. The present study examined the issue of how processing differed for these item types in the context of a problem-space framework. Protocols of examinees solving a small set of problems aloud were collected. These protocols were examined with respect to two phases of the problem-solving process: problem representation and problem solution. For formal-deductive items, all the information necessary to solve the problem was provided in the problem statement. The representation of formal-deductive items involved the use of meaning-reduced tokens and spatial diagrams. The units involved in the representation of informal reasoning and verbal reasoning item types included meaningful propositions and meaning-emphasizing paraphrases. Reference to common background knowledge occurred. The analysis of the problem-solution phase focused on the processes of evaluation, (judgments of the correctness of an option) and justification (statements of an argument or of evidence for why an option was or was not correct). First, the order of these processes was found to differ for formal-deductive items and other item types. Secondly, item solutions varied in terms of the kinds of justifications that were offered by the examinees for accepting or rejecting options. These results illustrate how the addition of some item types to the GRE analytical measure will expand the variety of reasoning skills assessed. Implications of these results for cognitive models of reasoning are also discussed.

## Acknowledgments

This study was supported by a grant from the Graduate Record Examinations Board. We wish to thank Margaret L. Redman for recruiting study participants and collecting data. Ruth Yoder's help in preparing the manuscript is greatly appreciated. We also wish to thank Randy Bennett, Clark Chalifour, and Kevin Singley for their helpful comments on an earlier version of this report.

## A Cognitive Analysis of Solutions for Verbal, Informal, and Formal-Deductive Reasoning Problems

In a recent study, Emmerich, Enright, Tucker, and Rock (1991) developed and pilot-tested additional item types for the GRE analytical measure, a test of reasoning, for the purpose of increasing the unity of the measure. Instead, they found that the additional item types loaded on three of four separable factors that were labeled verbal reasoning, informal reasoning, formal-deductive reasoning, and quantitative reasoning. The goal of the present study was to conduct a comparative, cognitive analysis of the types of reasoning items that loaded on different factors in the previous study. Such an analysis would help clarify why such a factor structure was found and would have implications for the validity of the GRE analytical measure as well as for psychological theories of reasoning. Below, we summarize the history of the GRE analytical measure and briefly describe a framework for the comparative analysis of reasoning items.

### Background and Related Research on the GRE Analytical Measure

In 1977 a test of reasoning, the analytical ability measure, was added to the GRE General Test. This measure was introduced to expand the range of reasoning skills assessed beyond those evaluated by the existing verbal and quantitative measures. Originally, the analytical measure included four types of items. However, two of these item types were found to be affected both by special test preparation (Powers & Swinton, 1984; Swinton & Powers, 1983) and by within-test practice effects (Kingston & Dorans, 1982; Swinton, Wild, & Wallmark, 1982). These two item types were eliminated from the measure in 1981.

Thus, from 1981 until the present, the GRE analytical measure has included only two item types: logical reasoning (LR) items and analytical reasoning (AR) items (see Table 1 for examples). Logical reasoning items consist of a short verbal argument followed by a single question or a pair of questions assessing any one of a variety of critical reasoning skills, such as recognizing assumptions, analyzing evidence, or drawing conclusions. Analytical reasoning items include a brief scenario and a set of rules about how elements in the scenario can be combined, followed by a set of questions. The analytical reasoning item type emphasizes deductive reasoning skills. Problems with the convergent and discriminant validity of this narrowed measure have been noted, however. Logical reasoning items correlate more highly with verbal items than with analytical reasoning items, and analytical reasoning items correlate better with quantitative items than with logical reasoning items (Wilson, 1985). Other studies using full information factor analysis (Schaeffer & Kingston, 1988) and confirmatory multidimensional item response theory (Kingston & McKinley, 1988) indicate a weak analytical factor defined by analytical reasoning items but not logical reasoning items. Finally, Rock, Bennett, and Jirele (1988) found that a four-factor solution with logical reasoning and analytical reasoning items constrained to load on separate factors fit better than a three-factor model.

Because the original analytical measure, with four item types, yielded a reasoning factor that was more distinct from the verbal and quantitative factors (Powers & Swinton, 1981), a study was conducted to develop more item types for the analytical measure in the hope of improving its convergent and

discriminant validity. Emmerich, Enright, Rock, and Tucker (1991) developed four additional item types for the analytical measure and evaluated them. The four additional item types are described below; examples of each types are included in Table 1.

Analysis of Explanations (AX) (revised). This item type, developed by C. Tucker, is based on C. S. Peirce's (Hartshorne & Weiss, 1931-1958, 2.776-7, 6.469) ideas about abduction (that is, hypothesis formation). A situation is described in a passage and a result is stated that appears paradoxical and in need of explanation. The examinee is then asked to consider each of several statements. For some statements, the examinee is asked to decide whether the statement is or is not relevant to any possible adequate explanation of the result. For other statements, the examinee is asked to judge whether the statement could adequately explain the result. An earlier version of this item type, with a fixed response format (answer options were the same for all items), had been included in the original analytical measure. However it was dropped because of apparent practice and coaching effects (Swinton & Powers, 1983; Swinton, et al., 1982) possibly related to the fixed response format. In the Emmerich et al. (1991) study, the item type was revised to include options unique to each item.

Numerical Logical Reasoning (NLR). This item type was based on work by Ward, Carlson, and Woisetschlaeger (1983) who attempted to develop "ill-structured" problems in a multiple-choice format. "Well-structured" problems are deductive in nature and require the manipulation of symbols as tokens and the application of algorithms. "Ill-structured" problems are complex, do not have definite criteria for determining when a problem is solved, and lack some of the information needed to solve the problem (Simon, 1978). In the "ill-structured" problems developed by Ward et al., the stimulus material is presented in the form of a chart, graph, or table. The question asks the examinee to analyze or evaluate a stated finding and/or other information in the table. For example, two contrasting interpretations of the data might be presented and the examinee asked to select the option that best supports one of those interpretations. As another example, the examinee may be asked to select the best explanation for the data.

Contrasting Views (CV). This item type presents two contrasting views and then asks a series of questions bearing on both views. Each view centers on a concept, that is expressed by the same word in each view, but that nevertheless differs in its implications in each view. Thus, the two views can be seen as alternative interpretations of the concept. Some of the questions measure the ability to recognize common aspects (central concepts or common assumptions), whereas others focus on aspects of disagreement (differences in implications or interpretation). Still other questions measure the ability to determine the relationship of a third view to the two given views.

Pattern Identification (PI). This is a form of "number series" problem. A sequence of numbers is presented to examinees who are required to select, from a set of answer options, another sequence of numbers whose pattern matches that embodied in the first sequence. In this approach, the task of formulating an applicable series rule is left to the examinee. However, to

ensure that the correct answer is unique and defensible, constraints are placed on the permissible operations in formulating the rules that governed the series. The permissible operations are limited to addition, subtraction, multiplication, and division of positive integers less than or equal to 3. (This item type was not included in the present study or in Table 1 because there were indications of a practice effect in the previous study, possibly related to the complex directions accompanying this item type.)

To evaluate these new item types, Emmerich et al. (1991) administered an experimental battery including the new item types as well as the two item types currently on the analytical measure to a sample of approximately 370 examinees. Data from an administration of the GRE General Test in December 1988 were also available for these examinees. The items on the experimental battery were administered in a 3-option format rather than the 5-option format currently used on the General Test; this was done to determine whether test efficiency could be improved. The results of an exploratory factor analysis performed on the data from the regular GRE General Test and from the experimental battery suggested that the reasoning domain could be divided into two subdomains (see Table 2a). The factor analysis was performed on parcels of from 4 to 15 same-type items from the GRE General Test and from the experimental battery (Cattell & Burdsal, 1975). Using Promax, four factors (principal components) for which the eigenvalues were greater than 1.00 were rotated. The resulting factor loadings are presented in Table 2a and the interfactor correlations as well as the variance explained by each of the four factors are presented in Table 2b.

The verbal and quantitative factors identified in the exploratory factor analysis represent dimensions of verbal and quantitative knowledge as well as some aspects of reasoning involved in comprehension. As seen in Table 2a, the least complex verbal item types (antonyms, analogies), which may be viewed as in part assessing verbal knowledge, best define the verbal factor whereas data interpretation, a quantitative item type that assesses individuals' abilities to understand the meaning and implications of quantitative information presented in tables or figures, best defines the quantitative factor. Thus, these two factors have a strong component related to specialized declarative knowledge. However, more complex kinds of items that require reasoning, such as reading comprehension and discrete quantitative items, also load on these factors.

The other two factors we identified represent separable dimensions of reasoning and have been labeled informal reasoning and formal-deductive reasoning. Note that the two item types, LR and AR, that compose the current GRE analytical measure load on different factors. Similar distinctions among modes of reasoning have emerged in the fields of philosophy (Toulmin, 1958), education (Voss, Perkins, & Segal, 1991), and cognitive psychology (Galotti, 1989). For example, in a recent review of psychological research on reasoning, Galotti (1989) distinguishes between formal reasoning and everyday or informal reasoning. She defines as a critical feature of formal reasoning that all the information to be considered is explicitly set forth in a problem. On the other hand, informal or everyday reasoning requires a search for relevant information or the determination of what information is relevant to the problem under consideration. In another discussion of this

distinction, Voss, Blais, Means, Greene, and Ahwesh (1989) note that both formal and informal reasoning center around the evaluation of arguments. In formal reasoning, the processes typically involved include converting propositions to symbolic form, combining these propositions to deduce new information, and determining whether symbolic relationships are in accord with the rules of the system. This is the kind of reasoning applied in formal deductive systems such as logic and mathematics. On the other hand, "informal" arguments consist of conclusions or hypotheses supported by reasons and are evaluated in terms of their soundness. "Informal" here does not connote carelessness in reasoning. Rather, informal reasoning employs different criteria from formal reasoning, which is concerned with validity and consistency rather than with relevance and consonance with a body of background knowledge. Among the processes involved in assessing the soundness of nondeductive arguments are evaluation of whether information is relevant to conclusions, whether and to what degree information supports a conclusion, and whether all relevant information that could support an alternative conclusion has been taken into account.

In examining the types of items that characterize the two dimensions of reasoning illustrated in Table 2a, this distinction between informal and formal-deductive reasoning seems very appropriate. Most of the item types that load on the informal reasoning factor (logical reasoning, numerical logical reasoning, analysis of explanations) have a stimulus that includes a number of propositions and a result, conclusion, explanation, or interpretation. The probes often require the examinee to determine whether additional information is relevant to a conclusion or to explaining a result or whether the information weakens or supports specific conclusions or interpretations. On the other hand, the formal-deductive factor is defined by the analytical reasoning item type, which requires examinees to deduce information from a set of conditions, and by mathematical item types from the quantitative section of the test.

The degree to which these types of reasoning draw on similar or different cognitive processes is a matter of debate at this time (see summary by Galotti, 1989). The results of our factor analysis support the view that these dimensions of reasoning call on different processes.

## A Framework for the Comparative Study of Reasoning Problems

Many psychologists who have studied "reasoning" have focused primarily on formal-deductive problems such as syllogisms (Braine, 1978; Johnson-Laird, 1983). A central issue has been whether or not people reason on the basis of a formal logical system. An alternative view is that reasoning can be understood in terms of the kinds of information-processing operations that are used to describe other forms of cognitive activity (Gellaty, 1989). Galotti (1989) described three different programmatic approaches to the study of reasoning that have emerged in the past 20 years: the componential approach, the rule/heuristics approach, and the mental models/search or problem-space approach. A problem-space approach is a particularly appropriate candidate as a framework for the comparative study of different kinds of reasoning problems because it is flexible enough and complex enough to allow contrasts among reasoning problems that vary in a number of ways.

In research on problem solving, individuals' representations of problems are analyzed in terms of a problem space, which includes a statement of the elements in the problem, the operations that can be performed on the elements, the goal of the problem, the constraints operating in a situation, and strategies useful in solving the problem (Greeno & Simon, 1988). Initially, research in this area focused on well-structured or knowledge-lean problems (Greeno & Simon, 1988; Reimann & Chi, 1989). In well-structured problems, solutions are governed by a system of logic such that the correctness of an answer can be demonstrated unambiguously or proved within that system. The role of semantic factors is minimal. Problem solvers operate on a set of objects, symbols, or tokens that are abstracted, to some degree, from the semantic context. The analytical reasoning item type currently used on the GRE analytical measure provides a good example of well-structured or knowledge-lean problems in which the problem elements and permissible operations are specified. Research on well-structured problems often has been concerned with describing how novices and experts differ in their representations of a problem and their solution strategies (see Reimann & Chi, 1989, for a recent summary).

In other research, the processes involved in solving ill-structured problems, which are often knowledge-rich, have been explored. Some of the characteristics of ill-structured problems are that they do not have formal criteria for determining when a problem is solved, and that established procedures for solving the problems do not exist. Furthermore, ill-structured problems may have a number of good, alternative solutions (Galotti, 1989; Simon, 1978). Solving knowledge-rich problems, such as diagnosing an illness, differs from solving knowledge-lean problems in two ways (Reimann & Chi, 1989). First, problem representation is not simply a matter of abstraction. The problem solver has to bring background knowledge to bear in developing the problem representation. Second, operators used may be more domain specific, for example, recognizing symptoms of a particular disease or knowing how to transform an algebraic equation. Many of the item types that load on the verbal reasoning and the informal reasoning factors in Table 1 have more in common with ill-structured problems than with well-structured ones because elements and operations are not clearly specified.

Investigators who have used a problem-space framework to study reasoning on verbally complex tasks have found it necessary to expand the framework in two important ways. First, it has been suggested that the problem situation is comprehended by developing a "situation model" or representation that draws on background knowledge as well as problem-specific information concerning the concepts, events, persons, or actions involved in the situation (cf. Groen & Patel, 1988; Hall, Kibler, Wenger, & Truxaw, 1989). Second, a need to describe the "reasoning structure" characteristic of the arguments that problem solvers offer to justify proposed solutions and to supplement more traditional descriptions of problem-solving control structure has been recognized (Voss, Greene, Post, & Penner, 1983).

Thus, research within the problem-space tradition on both well-structured, knowledge-lean problems and on ill-structured, knowledge-rich problems indicates a number of ways in which processing of GRE items that load on different factors might differ. These include differences in both problem-

representation and problem-solution activities. In the study described below, we sought evidence of how item types meant to assess reasoning varied in aspects of problem representation and problem solution.

## Method

The goal of this study was to conduct a comparative analysis of different types of reasoning problems that had loaded on different factors in our previous research. The basis for this comparative analysis was a description of some features of the problem-solving process that differentiated the way competent reasoners solved different types of reasoning problems. "Competent" reasoners, as defined by good performance on the GRE analytical section, were selected so that difficulties in general problem-solving skill would not obscure differences related to problem type. Students who had previously taken the GRE General Test were recruited and asked to solve reasoning problems aloud. The students' problem-solving protocols were then examined to identify features that were characteristic of the problem-solving process for different kinds of problems.

### Participants

Participants were recruited from local colleges and universities through advertisements in college newspapers and fliers posted at various locations at the schools. Sixteen undergraduate and graduate students who had taken the GRE General Test within the previous four years, who had scored above 600 on the analytical section, and whose best language was English were identified and agreed to participate. One-half of the participants were majoring in the humanities or social sciences and the other half were majoring in the natural sciences, engineering, or mathematics. Within each of these two major-field groupings, half the students were female and half were male.

### Procedure

Participants were tested individually in two sessions, each lasting from one to two hours, that were scheduled approximately a week apart. During each session students completed a paper-and-pencil test consisting of items of a particular type, and then concurrent verbal protocols (Ericsson & Simon, 1984) were recorded while they worked a different set of items of the same type aloud. Two or three of the five item types were administered during each session, and the order of administration was varied from participant to participant. Participants worked alone in a room and an experimenter was present in an adjacent office. After completing the paper-and-pencil test, administered in order to familiarize the students with a particular item type, the participants were asked to talk aloud as they solved each problem. More specifically, they were asked to summarize the initial statement of the problem situation or argument and then to talk aloud as they read the questions and answer options and to say what they were thinking. The session was videotaped with a camera positioned so that any marks, notes, or diagrams made on the test booklet were recorded. Typed transcriptions of the think-aloud protocols were prepared. Participants were paid $75 for completing the two sessions.

## Items

The materials consisted of some of the 3-option multiple-choice reasoning problems included in our previous study of new problem types for the GRE analytical section. The five types of reasoning problems examined in the current study included contrasting views, analysis of explanations, logical reasoning, numerical logical reasoning, and analytical reasoning. (As noted earlier, the pattern identification item type was not included because there were indications of a practice effect in the previous study.) For each problem type, a portion of the problems were presented in a paper-and-pencil format and the remaining problems were used in the "think aloud" procedure described above. Those problems that are the subject of the analyses in this study are presented in Table 1.

## Results

To illustrate contrasts in the reasoning applied to different types of items, we first present some protocol excerpts and discuss them in relation to a problem-space framework. Then selected features of the protocols are analyzed in more detail with respect to how problem representation and problem solution activities differed for formal-deductive, informal, and verbal item types.

### A Preliminary Analysis of Some Protocols within a Problem-Space Framework

In the following sections, we consider some examples of examinees' problem-solving protocols within a generalized problem-space framework. This approach illustrates areas of contrast among problem solutions for different types of items and motivates more detailed analysis of selected features of the protocols in subsequent sections.

Analytical Reasoning. An analytical reasoning stimulus typically begins with a scenario that defines which aspects of the entities it mentions are to be relevant for the task:

> AR94-98: An airline company is offering a particular group of people two package tours involving eight European cities--London, Madrid, Naples, Oslo, Paris, Rome, Stockholm, and Trieste. While half the group goes on tour 1 to visit five of the cities, the other half will go on tour 2 to visit the other three cities. The group must select the cities to be included in each tour. The selection must conform to the following restrictions:

The scenario typically presents a problem that is being addressed (selecting the cities to be included in each tour) and one or more lists of entities (the cities) that will be manipulated in the solution. The scenario is not fully formalized, and elements of common background knowledge (the way tours are understood as including cities) can play a part in establishing the relationships that are important to the task. Nevertheless, reference to subject-matter knowledge is sharply restricted, and essential relationships are clearly spelled out. The entities that are to be manipulated appear primarily as labels, or tokens.

Next, a set of rules for performing the required manipulations is typically given:

AR94-98: Madrid cannot be in the same tour as Oslo.
Naples must be in the same tour as Rome.
If tour 1 includes Paris, it must also include London.
If tour 2 includes Stockholm, it cannot include Madrid.

These rules are even more sharply restricted than the scenario with respect to reference to subject-matter knowledge. They provide a set of conditions analogous to axioms in a logical system, whereas the entities listed are analogous to logical constants in such a system. There are no explicit operations specified for manipulating the entities to solve the problem, and a natural-language interpretation of if, and, or, not, every, some, and so on, along with any deduction schemata such as modus ponens (if you have p, and you have p implies q, then you have q), are to be supplied by the problem solver. The task is to prove the correctness or incorrectness of answer choices on the basis of the given rules or the given rules plus information given in the question stem, such as "If tour 2 includes Rome...."

This proof process may be characterized in terms of Newell and Simon's (1972) "problem-space" description of problem-solving behavior. Each item may be seen as consisting of a set of states, including a start state and a goal state. Operators for transforming one state into another are not explicitly given, however, and evaluation criteria for states are also not fully explicit. Thus, analytical reasoning items are not fully specified, either as a logical system or as a problem-solving process. Still, all the information on which the implicit operators and evaluation criteria operate is given, and the material can be regarded as partially formalized.

The following protocol of a solution for an analytical reasoning item illustrates how an examinee's behavior may be characterized within a problem-space framework. First, this examinee reads and summarizes the problem stimulus:

AR94-98: An airline company is offering a particular group of people a two package tours involving eight European cities--London, Madrid, Naples, Oslo, Paris, Rome, Stockholm, and what I will pronounce Trieste. While half the group goes on tour #1 to visit five of the cities, so tour #1-five cities, that's one-half of the group, the other half goes on tour 2 to visit the other three. The group must select cities from the following--from, to be included in the tours. They must conform to the following restrictions:

M not O, that goes in both directions. N plus R. Tour 1 P then L. Tour 2 Stockholm not Madrid.

These comments are accompanied by notations in the examination booklet. The first letters of the cities are listed. A table-like notation is used to indicate the number of cities in each tour:

| $1_T$ | $2_T$ |
|---|---|
| 5 cities | 3 cities |

The restrictions described in the stimulus, for example, "If tour 2 includes Paris, it must also include London," are not read verbatim but are abbreviated both verbally and in writing in the booklet, for example, $2_T$ (S not M). Thus, as the examinee reads the stimulus, a representation of the problem is formulated that consists of a list of elements, rules, and a framework for manipulating the elements.

The examinee's comments as one of the problems in this AR set was solved were as follows:

AR95. If tour 2 includes Rome,
well then it must include Naples, we know that.

Which cannot be true?
Okay let's go through these.
So let's go, 2 then 1.
Well if tour 2 includes Rome then it must also include Naples.
Okay.
Tour 1 would have Paris and London.
That's fine.
Okay.

Here, in the problem-solution phase, the examinee sets up a table and invokes and applies rules even as the stem question is being read. The stem question describes the initial state, Rome in tour 2, and the examinee applies two rules and constructs a table with "P" and "L" in tour 1 and "R" and "N" in tour 2.

Finally, the options are read and commented on.

(A). Can Trieste be in tour 1.
Ah! Wait a second,

(B). if Madrid is in tour 2,
that's okay.
Madrid and Oslo can't be in the same tour.
So Madrid must be in one and Oslo must be in the other.
Okay.

(C). Stockholm is in tour 2
that's okay,
cause if Stockholm's here, Madrid's here.
Ah! but if Stockholm's here in tour 2, doesn't have Madrid, Madrid is in 1 but Madrid is stuck with Oslo.

So that cannot be true.
If S, right, well M can't be in it.
M isn't in it then M and Oslo.
So that can't be true.
Okay looking for (C).

Option (A) is not really considered. For both option (B) and option (C), the examinee first offers an evaluation of the option, "that's okay." The basis for this evaluation is not stated. (Offering an evaluation of an AR option prior to working out a proof is unusual, as we document in a later section.) Then selected rules are applied and the legality of the resulting state is evaluated.

Note that problem solution does not proceed by attempting to generate all possible consequences before comparing answer choices to them; there are, strictly speaking, an infinite number of possible consequences, and proofs would be inefficient if not directed toward a goal. Instead, the flow of processing is directed by the answer choices themselves. Processing requires steps of active inference that go beyond checking against the given rules. To show that (C) "Stockholm is in tour 2" is impossible, given Rome in tour 2, the inference is made that Rome entails Naples in tour 2. With Stockholm in tour 2, tour 2 would then be complete, so the other cities, including both Madrid and Oslo, would have to be in tour 1; but Madrid is not permitted to be in a tour with Oslo. This is an example of determining consistency by reductio ad absurdum: setting up a hypothesis (Stockholm in tour 2) and then deriving a contradiction with given information (Madrid is/is not in a tour with Oslo), thus proving the hypothesis impossible, or not consistent with the given information. In this item determining consistency or inconsistency is done by step-by-step inference, manipulating the tokens (city names), and not by making a global judgment.

Analysis of Explanations. In contrast with the AR problem described above, it is more difficult to characterize the solution to informal reasoning problems within problem-space framework. The following protocol for an AX problem illustrates some of the problems encountered.

One examinee's summary of the AX problem stimulus was as follows:

AX86-89: Okay. So we have a situation where a woman decides not to run after being in the state legislature for two terms. She would, knew it would be tough to find a job related to politics that would give her sufficient income and time to write. Uh, since she left college she'd been involved, so her background is all in politics. Um, she was popular. She was likely to win if she ran again. And she also was concerned about the probability of her party hurting because she wouldn't run. She found out that all these things she could put aside because she could find a job that they were willing to give her. And, I wonder if she made enough money there, and had enough time to write? Okay, and Louise Jones, a highly-qualified candidate, was willing to run in her place, so it shouldn't hurt her party that she was backing out. So her ego may have kicked in, apparently, so she decided

well, she wants to go back just because she doesn't want anyone else taking her place or if she's still not making enough money, so regardless of Louise Jones she still wants to run.

In contrast with the summary of the AR stimulus, this summary of the AX stimulus is primarily a paraphrase of the statements in the AX stimulus and is not accompanied by any note taking or diagramming. Although the examinee elaborates the situation by speculating about reasons why Joan Deeker may have changed her mind, little evidence of the nature of the examinee's problem representation is apparent in the summary.

Note that the examinee does not make any comments after reading a stem question but, instead, immediately begins to evaluate and process the options.

AX87. Which of the following statements, if true, is relevant to some possible adequate explanation of the result?

(A) Deeker's first campaign for a seat in the state legislature was unsuccessful.

No.
She's talking, she knew she'd get elected, she will, but she didn't want to hurt her party, wanted to make sure someone would fill her shoes, wanted a job that could pay, first campaign is irrelevant.

(B) City in which the university is located is a considerable distance from the state capital.

Unless she wanted to keep her hand in, then perhaps she wouldn't want to take that job and decided to run again.

(C) Organization of teachers sent an investigating committee to look into new charges that the university's policies governing academic freedom were repressive.

Well, if she wanted to get a job related to politics that would provide sufficient income and time to write, and the job couldn't stand up to her, what was it, liberal social, she was involved with liberal social programs, therefore, she probably had views in terms of freedoms and rights, and if there was promise of the academic policies at the university, she might not want to accept the position there, so she would run again.

After reading option (A), the examinee offers an immediate evaluation and then offers a justification supporting the evaluation. For options (B) and (C), justifications are offered but evaluation is implicit. The nature of these justifications differs for each option and from the type of justification evident for the AR problem above. The justifications for the AR options above consists of step-by-step inference that is warranted by the given rules. However, the justifications for the AX problem consist of (a) a list of factors that are relevant to Joan Deeker's decision, (b) generation of

a hypothetical circumstance under which option (B) might be relevant, and (c) a chain of informal inferences relating information from the stimulus situation and option (C) to a plausible explanation. In terms of a problem-space framework, the sequence of propositions offered by the examinee can be seen as analogous to moves from one state to another, but the basis for such transitions is rooted in knowledge and experience rather then formal rules and is not always transparent.

Contrasting Views. The stimulus for contrasting views items consists of two juxtaposed views centering around a given concept, expressed by the same word in each view, but differently interpreted in each, so that implications and assumptions differ in the two views.

CV26-29:

18th-century view: The new science will liberate the human mind and provide us with a mastery of nature, with which we will break the bonds of tyranny, transform society, and improve all the conditions of life. Rank and birth will fall into contempt in the new age of democratic progress; science is progressive.

20th-century view: Science and technology make possible, not only new products from natural resources, but also new processes of production; not only new techniques of farming, but also new crops. This enables our industry and agriculture to remain competitive. Technical advances will unavoidably result in unemployment and dislocations of the industrial and farm labor force in our society; this is, however, the price of progress.

One examinee summarized the above as follows:

> Okay, the 18th-century view: Science is good and doesn't, this view does not approach any of the technologies or any science, per se; simply its effect on society and political concerns almost; while the second one discusses direct relationships between science, technology, and any societal impact. In the first, societal impact is all positive. In the second, there will be unavoidable negative impact which, well the price of progress--they seem to say progress is the positive and there will be a negative price for progress.

In this summary, the examinee proceeds to extract the main point of each view and to contrast the views with respect to more detailed points of similarity and difference.

The examinee's responses for one of the items associated with this stimulus follows:

CV27. Eighteenth-century view, but not the twentieth-century view, rests on the assumption that

(A) science is value-free and can be used either for good ends or bad.

No.
It talks only about the good.

(B) the privileged would invest in technology and world reap the rewards.

No.
It talks about the break in the tyranny and rank and birth will fall into contempt.
So rank is not important.
So privileged would not be affected, will be affected, and they would not be the only ones to reap the benefits.

(C) human power, (C), over nature would be used to benefit people who had held little political power.

Well then, you certainly will break the idea of rank, rank and birth.
So that's (C).

In this item the two views are evaluated with respect to their consistency with particular assumptions. The first option is rejected as inconsistent with the eighteenth-century view based on the main point of that view. The second and third options are rejected and accepted respectively because of inconsistency/consistency with specific points embodied in the eighteenth-century view.

The processing is not step-by-step, as in analytical reasoning, but includes global judgments of consistency that are based on the main point of each view as well as consistency judgments with respect to more specific details.

Our preliminary review of the examinees' protocols suggested a number of areas in which solutions for verbal, informal, and formal-deductive reasoning problems can be contrasted. First, the representation of the problem situation varied for different types of items. Secondly, the interaction between the evaluation of an option and its justification also seemed to differ among item types. A third area of contrast was the kind of justification offered as a basis for accepting or rejecting options. Each of these topics is considered in more detail below.

Problem Representation

Two particular aspects of problem representation can illuminate the contrast between reasoning processes; the nature of the units that problem-

solvers manipulate, and the frameworks within which these units are manipulated. The generation of these units and frameworks is largely silent, except where we can observe a diagram being developed, but evidence that such a process has occurred can be gleaned by examining the ways in which problem-solvers refer to these units and frameworks. Especially useful are instances in which references are made by pronouns such as "it" or "that" and adverbs such as "here." In the following sections we describe differences in the units manipulated and the frameworks developed in representing and solving formal-deductive, and informal and verbal reasoning problems.

Units manipulated: Formal-Deductive Reasoning. The primary units manipulated in analytical reasoning are generally proper names or labels making the units into tokens or counters, which can be moved according to the given rules as if on a game board. For the analytical reasoning questions on the tour package, nearly half of the examinees made an initial list of the first letters of the cities in the tour package, and all but one examinee frequently used letters rather than city names in their tables or diagrams. When problem solvers summarize a list of such units, they focus not on their semantic meaning (something like, "They are all capital cities except Naples and Trieste") but on their characteristics as symbols:

> AR94-98: London, Madrid, Naples, Oslo, Paris, Rome, S, T: L, M, N, O, S, T; L, M, N, O, P, R, S, T--where's the Q?

Even before the list has initially been read through, symbolization by letters has occurred, and then the problem solver comments on the sequence of letters as labels. What is occurring is that the names or labels are being treated extensionally, as logicians say. They do not have meaning outside their functioning, within the given system of rules, as distinguishable labels for different entities; put another way, their meaning is given in their specified relationships.

Complex units, such as pairs of primary units, or possibilities (sets of primary units that are consistent with the rules) are also generated in the course of problem solving. Beginning with

> AR94-98: Then it says Naples must be in the same tour as Rome. So, you could have Naples and Rome here or Naples and Rome here.

one problem solver progressed to calling the two cities (AR97) "the Rome-Naples double thing." Another said, (AR94-98) "So, Naples and Rome get stuck together." Treating the pair as a unit has the advantage of making explicit how much room is taken up within the tour groups (of three and five cities) whenever one of the pair is included. It incorporates the import of one of the rules into the unit manipulated, so that the step of checking against that rule for each solution is eliminated. Another problem solver attempted an exhaustive listing of the possibilities permitted by the rules but found the procedure difficult to control: (AR98) "That was hard."

When a rule is summarized, we often find that it is encoded in some sort of symbolism, such as R<->N for the rule about Rome and Naples or $1_T$(P then L) for the rule about Paris and London. Sixty-three percent of the examinees

encoded the rules in some such symbolic notation. We do not find a hypothesis generated about the reason or rationale for the rule, such as, "Madrid cannot be in the same tour as Oslo; that could be because they are so far from each other." Such rationales, incorporating background information, are characteristic of AX and NLR. Rather, we find a comment such as

> AR94-98: And it says that Madrid cannot be in the same tour as Oslo. So if you had Madrid here, so Madrid has to be in one and Oslo has to be in the other.

The rule is being treated as a given with no further ground or explanation, and the effort of comprehension is directed toward understanding the range of its implications.

Here forward reasoning has been done to reach a conclusion about a constraint on possible solutions, and this constraint is assimilated into the diagram or structure that is being set up. Other problem solvers, less efficiently, do not draw the conclusion that one of the places on each tour is taken up either by Madrid or by Oslo. Then they must check each solution against the rule. Later, through habituation, they may come to comprehend the structure required by the rule:

> AR95, second time through: If Stockholm were in 2, Madrid would be in the first, and Oslo would have to be in the second, that would be too many. Thank you! It's (C).

In addition to comments about the positive implications of rules, comments are made concerning what rules do <u>not</u> say, as in

> AR94-98: If tour 1 includes Paris, it must also include London. But that's not necessarily the other way around.

Here it is the rule that is treated as a unit (<u>that's</u> not necessarily). The problem solver is explicitly guarding against the common logical mistake of "affirming the consequent," or interpreting an ordinary "if" as "if and only if."

Going further in treating rules as units, one of the problem solvers gives the rules explicit labels (1, 2, 3, 4) and cites them by those labels in justifications:

> AR94: If I label these 1, 2, 3, and 4, by 3 and 4, Stockholm has [tour] 2, Madrid can't be in it [tour 2]. If Stockholm, then all of these [answer choices] are right by 3 and 4. So 1 and 2 are gonna be the ones [rules] that change it [the answer].

This attempt at higher level formalization, however, soon falls away as the line of thought fails to lead to a satisfactory answer, and the problem solver returns to citing the rules in a more active form, rather than by a label.

The most common way of referring to a rule is to repeat it, in whole or in part. Rules are also frequently referred to by paraphrase, such as (AR98)

"Naples would have to have Rome." It is rules or, rarely, parts of the introductory scenario that are paraphrased, and this transformation treats them as whole propositions having semantic meaning that can be recast in other words.

In analytical reasoning questions, the given material is not fully formalized, of course. The initial scenario does rely on the meanings of propositions and terms to establish the relevant features of the structure and units with which the problem solver will work. The description of cities as included in tours makes it clear, for instance, that the problem solver need not consider the possibility that the same city could occur twice within a tour (two visits) or that it could occur in both tours. The description also establishes that the tours together include all of the cities, not a selection from among the cities. Problem solvers behave accordingly without comment.

The material in the initial scenario is assimilated as background information, as is evidenced by problem solvers who have greater difficulty accessing information given there than they do information given in the rules. Item 94 includes an answer choice, (C), that satisfies all the given rules but contravenes a stipulation in the scenario--that there are five cities in tour 1 and three in tour 2. Several problem solvers were unable to see what was wrong with (C) until they finally went back to the initial scenario.

> AR94: (A) and (C) both seem okay to me.....Um, oh, okay. Five are in tour 1, so it's gotta be (A). That took long.

One problem solver failed to reach the correct answer because this information from the scenario was not taken into account.

Thus, the units used in analytical reasoning problem-solving include (a) the primary units, which are extensional or meaning-reduced symbols or labels, and then, depending on the problem solver, (b) combinations of these primary units into complex units such as pairs or sets (possibilities), (c) notations encoding, as part of a developing spatial structure, information given by the rules about the primary units, (d) higher level formalizations of rules in terms of extensional labels, (e) paraphrases giving the meaning of rules, and, rarely, (f) paraphrases of information from the background situation. All these kinds of units, except (e) and (f), exhibit abstraction of information from the semantic context.

Framework: Formal-Deductive Reasoning. Besides the general background of information provided in the scenario, analytical reasoning problem solvers tend to construct an explicit spatial framework in which to manipulate the primary units. Diagrams are constructed even for analytical reasoning stimuli that are not concerned with spatial ordering or any other kind of ordering. The stimulus about cities in tours is about assignment to set membership, not about order relationships, yet all 16 problem solvers made use of spatially organized diagrams or lists in answering at least one of the AR questions. They often use spatial terminology, such as (AR94-98) "on one side...on the other," (AR95) "on this side of the equation," (AR95) "put London over here," (AR96) "in either place," (AR95) "separated," (AR96) "there's not room for both," (AR97) "there's no space."

Although reasoning with the aid of a spatial framework is practiced by all of the problem solvers, they do so to varying degrees. Some rely primarily on the meanings or entailments given by the rules:

> AR95: I don't see any restrictions on Trieste.
>
> AR95: Uh, Stockholm is in 2 [cannot be true], because that would entail Madrid and Oslo being in the, in 1.
>
> AR97: If tour 1 includes Paris and tour 2 includes Madrid, which of the following must be included in tour 2? Um, what does Paris imply? If tour 1 includes Paris, it must also include London. What does Madrid mean? Madrid cannot be in the same tour as Oslo.

When controlled by a strong sense of the relevance of rules to tasks, as in (AR96) "We know that Naples must be in the same tour as Rome. But that's not yet relevant," this largely nonspatial mode of reasoning produces efficient proofs. Otherwise, however, it leads to incomplete proofs (with only one of two relevant possibilities considered, for example) and to random and repetitious wandering among possibilities, arriving in confusion at dead ends:

> AR96: Okay, so that would be XXXXXX [unintelligible] again. Except we have mostly the same riddles we had last time. Let's try something different. See if anything else will work out.

Analogous to doing mathematical calculations "in your head," this nonspatial mode of reasoning depends on a clear memory of the path taken.

What the spatial framework does for the problem solver seems to be to provide a way of keeping track of the progress of reasoning, just as doing mathematical calculations by manipulating symbols spread out spatially on paper reduces the burden on memory and provides a way of checking accuracy. The spatial framework does not encode each reasoning step--the rules are generally not represented in it--but enables initial conditions and intermediate and final results to be recorded.

Thus, we find that solving analytical reasoning problems is a fluid, fallible process of creating, and sometimes abandoning, units with which to work and of maintaining a train of thought, more or less successfully, under the control of relevance. Embedded in this process, however, is the manipulation of discrete extensionally understood symbols according to given rules within a spatial framework; this is the more formalized part of the proof process.

Units manipulated: Verbal and Informal Reasoning. The reasoning item types that load on the verbal and informal-reasoning factors (LR, CV, AX, NLR) are more similar to each other with respect to units manipulated and background or framework within which manipulation takes place than they are to analytical reasoning. For the most part, summation of the stimulus paragraph for AX, CV, LR, or NLR did not make use of diagrams. In addition to verbatim restatement or close paraphrases of stimulus information, examinees elaborated their summaries with comments about the main point or gist of the stimulus,

with informal inferences, with possible explanations, and with background information. Examples of these kinds of elaborations are presented in Table 3. The notational tactics most frequently used for these item types typically involved the emphasis of certain portions of the stimulus through underlining, circling, or bracketing. A few instances of writing an interpretive or summary label next to a portion of a stimulus were noted for CV and NLR items (e.g., "capitalist," "Enlightenment," "cost of operations").

In contrast with the primary unit manipulated in analytical reasoning, the primary unit manipulated in verbal and informal reasoning problem-solving is the statement or proposition:

> LR27: That [a statement constituting an answer choice] sounds possible.
>
> LR54: That sounds good....
> Well, that also seems to be, to conform with the statement above....
> Doesn't seem to be relevant.
>
> LR53: Well there's nothing that's presented that, uh, suggests that.
>
> LR6: And it doesn't actually support the claim...

Sometimes, the stimulus as a whole is taken together as a unit, encompassing what "it" says:

> LR36: (B) The lie detector gives accurate results only when employed. No, it doesn't say only,...
>
> CV26: And for progress, it's kind of neat cause originally I was looking at it as a discussion of science: and now, I'm looking at it as a discussion of what progress is.

In contrasting views, each view can be treated as a unit, as well.

These units (propositions, and stimulus either as a whole or as two views) differ from the primary units of analytical reasoning problem-solving, first of all in their size. They are units that are themselves complex wholes. Second, they are understood as having semantic meaning, unlike the symbols primarily manipulated in analytical reasoning.

Sometimes, information in the stimulus is condensed and focused in a summarizing statement:

> LR36: A study of the use of the polygraph, or lie detector, found that when a trained examiner using approved questioning techniques gave the test, information from the lie detector was accurate in determining whether responses were truthful for 70 to 90 percent of the responses.
> Okay so there's accuracy with a trained professional.

This condensing process omits some information, showing that this information has been assimilated as background or as detail; it also highlights other information as important or relevant. The omitted information establishing the topic ("polygraph, or lie detector") does not receive further mention, though it might serve as a background influence on what is relevant. The omitted information about "70 to 90 percent" is cited elsewhere in a justification:

> LR36: [in refutation of "definitely lied"] No because it's 70 to 90 percent.

Thus, that omitted information has remained available for use as detail even though it is not part of the central summary.

This condensing process also paraphrases statements and phrases, transforming the information in other words: "trained examiner" becomes "trained professional." This paraphrasing process contrasts with the meaning-reducing labeling of units in analytical reasoning problem-solving, where "Stockholm" was taken extensionally, equivalent to the label "S." Instead, the background of meanings that the problem solver brings to the problem is called on to explicate and categorize the given information.

The effect of this problem solver's summary is to express a perspective on the subject matter that slants judgment toward the acceptance of polygraph results. The term "professional" is more favorable than "examiner" in the stimulus, and "there's accuracy" is more favorable than "were truthful for 70 to 90 percent" in the stimulus. This transformation could have been influenced by the problem solver's antecedent views or by given information assimilated as background information (the name given in the stimulus for the device, "lie detector," implies that lies <u>are</u> detected). This slant makes it more difficult for this problem solver to accept a conclusion that follows with logical certainty from "70 to 90 percent" but not from "there's accuracy," namely, that the "lie detector failed to give correct results in at least one out of ten instances":

> LR36: Umm, maybe.

The problem solver nevertheless, after rejecting the alternative answer choices, accomplishes the transition in perspective from the problem solver's own constructed summary to the implications of "70 to 90 percent":

> LR36: So, let's say the minimum is 1 out of 10 times where there's a mistake.

This conclusion itself is a meaning-emphasizing paraphrase of the conclusion as presented in the answer choice. It is now the positive-sounding information about "trained examiner using approved questioning techniques" that has become omitted.

Besides entire propositions, short phrases are also manipulated as units, both by quoting them, generally with meaning-emphasizing reduction of detail, and by paraphrasing them. A stimulus phrase, (LR54) "confident of a

diagnosis of acute illness" is integrated into a problem solver's summary statement. Using forward reasoning, the problem solver supplies a hypothesis to explain the link asserted in the stimulus:

> LR54: Because you can be more confident , I assume, of the diagnosis. so...[underline added]

In this example, "more" is supplied by the problem solver to link the first claim of the stimulus with the second claim, which contains "more," and the detail "of acute illness" is omitted. This detail recurs, in part, in a further explanation the problem solver generates to link the statement in an answer choice to the forward-reasoned link already supplied.

> LR54(A): Therefore they can be more confident of an illness. Okay.

A paraphrase of "costs more" from the stimulus constitutes the entire statement of a hypothesis the problem solver generates to link the statement in another answer choice to the forward-reasoned link already supplied.

> LR54(B): [quoting] several remedies at once....More expensive. Okay.

That is, using several remedies at once would be more expensive.

The links generated in these instances, in contrast to those for analytical reasoning problems, are not primarily links between symbols. Rather, they represent subject-matter connections supplied by the problem solver. The problem solver knows or thinks that a hospital's accountant would charge more if emergency-room personnel use "several remedies at once," so that care would be "More expensive."

Another problem solver explained the subject-matter information that must be supplied to choose (C) in item 54:

> LR54(C): In fact, one would think that if there was a smaller number of illnesses treated in the emergency room, that they would be much more efficient and cost effective when treating the smaller number than in a private office.

Two abstractions were created as units by another problem solver:

> LR54: So you have confidence and conservativeness are the two things which lead you to, uh, have the cheaper treatment.

That problem solver's version of the hypothesis generated to link the two claims in the stimulus is

> LR56: it goes confidence, conservative, cheaper.

Here the single-word units are used as labels for complexes of ideas. Such a label can be viewed as a limiting case of meaning-emphasizing paraphrase, designating a "head idea" standing for a whole. This use of labels for complexes of ideas is particularly characteristic of contrasting views; one problem solver repeatedly refers to the optimistic 18th-century view of progress as "up-up-up."

In analysis of explanations, noun phrases from the text serve the same function of labeling complex ideas:

> AX86: The previous legislation and most popular legislation is unrelated to her decision because her decision was based on two other factors.

Similarly, for NLR38, many examinees selectively focused on the increase in the price of diesel oil as a main causal factor to be integrated into a explanation: "This doesn't deal with this issue here about the price of diesel oil" and "if we can relate these possibilities back to the issue of the price of diesel oil."

Thus we find that the primary units manipulated in verbal and informal reasoning problem-solving are meaningful wholes, generally propositions. In addition, both propositions and shorter phrases are reworked into units through the process of meaning-emphasizing paraphrasing.

Framework: Verbal and Informal Reasoning. For the verbal and informal reasoning item types, there is very little that corresponds to the spatial framework constructed by most problem solvers doing analytical reasoning. For logical reasoning, analysis of explanations, and numerical logical reasoning, no diagrams and very few spatial words are used.

For one contrasting-views practice problem, however, a problem solver did generate a diagram that differs in type from those generated for analytical reasoning (See Figure 1). This diagram uses meaningful terms as units and connects them by lines representing semantic relationships (intellect and sentiment as falling under content, where content represents a head idea). It is combined with underlining in the passages that gives emphasis to the words expressing important concepts. Note that in view II, form is circled in the word formal, to indicate that this is where the major concept form is discussed. The difference in the main points of the two views is symbolized by the contrasting notations form>content and form<->content. This diagram is visually similar to those that student writers are often taught to construct when organizing an essay. It is a concept map, rather than a game board on which symbols are to be moved.

In sum, major differences in the tactics examinees used to represent formal-deductive items, verbal items, and informal reasoning items were apparent. Reductionist notational tactics were used to summarize AR items, whereas the preservation of meaning and the use of background knowledge to elaborate the problem situation were common for CV, LR, NLR, and AX items. Verbal items and informal reasoning items had more in common with each other than either did with AR items in terms of the tactics used to represent the

problem situation. The schematic representation of the AR item set typically consisted of a list of meaning-reduced elements or tokens, symbolic encodings of rules, and a diagram with slots to be filled in. Overt schematic representations of the problem situation were rare for the other types of items, where the problem summary was dominated by immediate recall of text. However, single-word labels evident in summaries were meaningful, and background knowledge was used to elaborate the situation described. In contrast to the tokens that were the primary units manipulated in AR, the primary units manipulated in the other item types were meaningful propositions and meaning-emphasizing paraphrases.

We now turn to consideration of some differences among item types that were evident during the problem-solution phase of processing.

Problem Solution

Two distinguishable aspects of protocols during problem solution included (a) evaluations--statements concerned with the plausibility of options as good answers and (b) justifications--explanations of why options were or were not good answers. We analyzed differences among item types (a) in the order in which evaluations and justifications were presented by examinees and (b) in the kinds of justifications examinees offered. Examples of the kinds of evaluations and justifications offered for different options are presented in Table 4.

Ordering of Evaluation and Justification. For analytical reasoning items, it could be expected that evaluation would follow justification, since the explicit rules given in the stimulus allow one to construct a deductive chain of inferences leading to a necessary conclusion. These inferences serve as a certain basis for evaluation of the validity of the option. For other types of reasoning problems, however, the order of these processes would seem to be less predictable. On the one hand, it is always advisable to think or reason before making an evaluation. On the other hand, informal judgments of relevance, consonance, or importance seem to have an immediacy that judgments of deductive validity do not. Therefore the examinees' comments about each option were classified into one of five categories described below.

1. None--no comment about the option.

2. Evaluation only--only a comment about the plausibility of an option is made, or the option is restated and related to the stem.

3. Justification only--a justification is offered but evaluation is implicit rather than explicit.

4. Evaluation followed by justification--both an explicit evaluation and a justification are present, with the evaluation offered prior to the justification.

5. Justification followed by evaluation--both an explicit evaluation and justification are present, with the justification offered prior to the evaluation.

Although examinees sometimes reevaluated options after justifying them or engaged in more than one episode of justification and evaluation, only the first justification/evaluation episode was coded for each option, since our primary interest was whether examinees offered an evaluation prior to a justification or vice versa. The proportion of examinee responses to options that were classified in each category for each item type are presented in Table 5. There are very consistent similarities in the pattern of responses among the CV, LR, NLR, and AX item types, and the pattern for these item types differs from the pattern characteristic of AR items. As expected for AR items, justification typically preceded evaluation (60% of the time), whereas evaluation preceded justification only 10% of the time. However, for the other item types, evaluation often preceded justification (from 31 to 42% of the time) but justification preceded evaluation only 10 to 14% of the time. Similarly, examinees were more likely to offer only an evaluation than to offer only a justification for CV, LR, NLR, and AX; the opposite was true for AR items. We suspect that this tendency for examinees to offer immediate evaluations for these item types is related to the integration of problem information into a problem representation that provides a basis for making rapid judgments of the relevance of other information.

Categorization of Justifications. A system for categorizing the justifications examinees offered was developed through an iterative process. Initially, two of the authors reviewed the protocols of four subjects and developed a preliminary categorization scheme for the subjects' responses to each option. This scheme was applied to a larger subset of the protocols by the first and third authors and refined through discussion of disagreements and ambiguities. Finally, the protocols for all of the item types except analytical reasoning were coded by two of the authors, and, after further clarification of the categories, disagreements were resolved by one of the authors. (The coding of analytical reasoning items was less ambiguous than that for other kinds of items, so not all such items were double coded.) The resulting scheme, which included 12 categories, is described in Table 4. Agreement on the categorization of responses to each option was 69% for a subset of four subjects. More than one category could be applied to the response to an option if two distinct justifications were thought to have been present. Over half the disagreements concerned whether a justification was actually presented or whether one or more justifications needed to be coded. The former situation frequently arose when a subject reiterated and combined the information from the stem and the option and evaluated the option but offered no additional support. Although for some items this is all that really is required to evaluate the option, such responses were classified conservatively as "no justification" in the final coding.

The proportion of justifications in each category for each item are presented in Table 6. P+, the proportion correct found when the problems were administered to a sample of 374 students in our previous study, is included in Table 6. The proportion of times that no rationale was offered or was unintelligible or unclassifiable for each item is also presented in Table 6. The most obvious pattern in this table is the extent of the overlap in the kinds of justifications offered for CV, LR, NLR, and AX item types. List justifications were frequently used for all of these item types. However, there are also some distinctions among these item types that may help to

account for the factor structure found in our previous study. The predominant justifications offered for the set of CV items examined in this study, which were included in parcels that loaded on the verbal reasoning factor, involved lists, interpretations, and generalizations. Informal inferences and suppositions were frequently used to justify NLR and AX items, which loaded on the informal reasoning factor and are typically concerned with explanation. Parcels of LR items are inconsistent in their factor loadings, loading variously on the verbal reasoning factor, the informal reasoning factor, and even on the formal-deductive factor. Differences in the pattern of justifications offered for the four LR items examined in this study suggest that LR items may be more heterogeneous than other item types. Although three of the items examined had a justification pattern similar to CV items, the other item, which was concerned with explanation, was more similar to NLR and AX items. Alternatively, the apparent heterogeneity of LR items may simply reflect the fact that each LR item in this study had a unique stimulus whereas CV, AX, and NLR items came in "sets" based on one stimulus situation. In the present study only one CV set, one AX set, and two NLR sets were used. If examples of CV, AX, and NLR items from different sets were compared, they could possibly exhibit more heterogeneity in terms of the justifications offered for their solutions. Finally, as we expected, there was little overlap between the kinds of justifications offered for AR items and other kinds of items. The inclusion of explicit rules in the stimulus statement for AR permits kinds of rule-based reasoning that are not commonly found for other item types. Furthermore, for some AR items, a small set of possible tour combinations can often be determined through rule application prior to any consideration of the options. Then the correct answer can be selected by matching options to the set of possible combinations. This type of "generate and test" strategy was common for AR95 and AR97.

## Summary and Discussion

The "two disciplines of scientific psychology" (Cronbach, 1957) have been integrated fruitfully in the past two decades in many studies. Snow and Lohman (1989) note a number of ways that cognitive psychology is contributing to progress in educational measurement, including improved construct validation, the development of alternative measurement strategies, and improved theories of aptitude, learning, and achievement. The interaction between cognitive theory and measurement research, however, can take the form of a dialogue. Performance on assessment tasks can be a stimulus to psychological theorizing, and the measurement field can provide a testing ground for psychological theories. Measurement results should provoke psychological theorizing, and better cognitive theories should provide principles that might improve test and item design.

In the current study, we use some unexpected findings from a measurement study to stimulate thinking about psychological models of reasoning. One of the most intriguing outcomes of the Emmerich et al. (1991) study, which sought to develop a more unified reasoning measure, was that the proposed item types actually loaded on three of four separable factors. This result illustrates how limited our psychological models of reasoning are in that they provide little guidance about what the factor structure among reasoning tasks is likely to be. For many years, cognitive psychologists have focused their

attention on formal-deductive kinds of problems. Recently, interests have broadened considerably, and topics such as explanation (Thagard, 1989), argumentation (Kuhn, 1991), induction (Holland, Holyoak, Nisbett, & Thagard, 1986), and informal reasoning (Voss et al., 1991) are now being studied in depth. Despite this broadened interest, the comparative study of different kinds of reasoning has seldom been carried out. The results of our empirical study of different types of reasoning problems (Emmerich et al., 1991) raised the issue of how processing might differ for problems that were psychometrically distinguishable.

We explored this issue in the context of a problem-space framework. Protocols of examinees solving this subset of problems aloud were collected. These protocols were examined with respect to two phases of the problem-solving process--problem representation and problem solution. For formal-deductive AR items, the representation consisted of meaning-reduced tokens, a spatial framework, and the rules given in the problem statement. For verbal reasoning and informal reasoning item types, the primary units manipulated were meaningful propositions and meaning-emphasizing paraphrases. A schematic framework representing the problem situation was generally absent.

The analysis of the problem-solving phase focused on the processes of evaluation (judgments of the correctness of an option), and justification (statements of an argument or of evidence for why an option was or was not correct). First, the order of these processes was found to differ for formal-deductive problems and other types of problems. Examinees often evaluated options for informal and verbal reasoning problems before offering a justification for an option, but justifications preceded option evaluation for formal-deductive problems. Secondly, items also varied in terms of the kinds of justifications that were offered by the examinees for accepting or rejecting options.

Implications for the Assessment of Reasoning

Perhaps the most important implications of this study for the GRE analytical measure concern construct validity. It is clear from the results of both this study and those of Emmerich et al. (1991) that adding new item types such as AX, NLR, and CV will broaden the range of reasoning skills assessed by the GRE General Test. AX, NLR, LR, and CV are distinct from AR in terms of factor structure, in the cognitive processes involved, and in the forms of argumentation used to justify an answer. Using a wider variety of item types that require different modes of reasoning will better represent displinary diversity in reasoning (Toulmin, Rieke, & Janik, 1984).

The results of this study help to clarify the contribution of different kinds of reasoning to the exploratory factor analysis reported in Emmerich et al. (1991). The item types that load on the verbal factor (CV, LR) often involve reasoning about meaning and interpretation. Some of the item types that load on the informal reasoning factor involve explanatory reasoning (NLR, AX, LR). These results suggests that items that vary in the required mode of reasoning load on different reasoning factors. However, this implication needs to be confirmed through additional study because the current results

were obtained from a small, select sample of examinees and a small set of items.

Other issues about the factor structure of the GRE General Test, and the characteristics of items that contribute to the factor structure also require further thought and investigation. For example, what combination of items will lead to a unified analytical measure? Emmerich et al. (1991) considered how different combinations of reasoning item types might affect the unity of the analytical measure. They concluded that although the unity of the measure would be greatest if it were composed of either formal-deductive or informal reasoning item types, some improvement in unity of the current measure would be gained by adding more of the item types under investigation. This improvement in unity would be due in part to the inclusion of the pattern identification item type (an inductive number-series problem type), which loaded on both the informal reasoning and formal-deductive factors and was not investigated in the current study. If the PI item type is not included, the outlook for improving the unity of the current measure is probably less positive, as our results indicate that CV, LR, NLR, and AX do not have a lot in common with AR in terms of the cognitive processes involved.

Furthermore, the conclusions of Emmerich et al. (1991) about improving the unity of the analytical measure were based on factor analyses in which parcels composed of items of a particular type were used. Our results concerning the nature of justifications indicate that there may be important differences and similarities within item types that need to be taken into account in test design. Some LR items may have more in common with verbal reasoning items than they do with informal reasoning items. This difference may parallel test development subcategories of LR items with respect to whether or not an item centers on the meaning of a term or on an explanation. The issue to be raised here is whether these subcategories of items load on the same or different factors. At a more general level, we need to document systematically what similarities and differences among items contribute to the correlational structure of the test. This issue is particularly important in the context of computer-adaptive tests, in which different examinees answer different items.

Implications for Models of Reasoning

Modeling reasoning on tasks that require extensive background knowledge presents an extremely challenging problem for cognitive scientists. One example of an attempt to do so is the work of Collins and Michalski (1989), who have developed a theory of plausible reasoning that includes a formal representation of plausible inference patterns that are evident in people's answers to everyday questions about the world. They note that one criticism that has been made about their approach is that verbal protocols do not expose nonverbal processes that may contribute to a problem solution; critics have pointed out that "verbal protocols may be rationalizations for answers arrived at by some other process (p. 41)." Our results concerning the frequency with which examinees evaluate options prior to explaining their reasoning partially supports this view. Although we agree with Collins and Michalski's response "that answers follow frequently from both verbal and nonverbal reasoning processes and that these are weighed together in responding (p. 41)," more

attention needs to be given by researchers to what the nature of these other, nonverbal reasoning processes might be. The problem of compiling a corpus of "common-sense" knowledge that permits understanding and rapid judgments of relevance has proved to be a difficult hurdle for researchers in the field of artificial intelligence, and some have concluded that the dominant paradigms based on encoding of rules and facts, or cumbersome proposition networks, will fail in the long run (Dreyfus, 1992). The emergence of connectionist models may offer a new avenue for modeling the preverbal processes that seem to be an important part of informal and verbal reasoning.

The rapidity of response and lack of articulated reasoning preceding evaluative judgments in informal reasoning needs explanation, and the absence of step-by-step processing may serve as a clue. Because problem solvers make many immediate or nearly immediate judgments in informal reasoning, the framework within which informal reasoning takes place is likely to be the stimulus material assimilated as a whole.

We can hypothesize that this whole is analogous to a field rather than to a network of discrete units that would have to be checked one by one. It would be meaning field established by the interaction of the meaning units in the stimulus with one another as well as with relevant background knowledge or opinion. Rapid recognition of fit or lack of fit between the smaller whole constituted by a single proposition and the larger whole constituted by the stimulus could well occur in terms of overall properties of the two semantic fields. Analogously, proteins, which are very complex molecules, are recognized quickly by the equally complex molecules in cell receptors by virtue of overall properties of shape, rather than by atom-by-atom checking.

On this hypothesis, processing in terms of units smaller than a proposition would tend to occur when it was insufficiently clear whether adequate fit had been achieved. Then processing in terms of highlighting and meaning-emphasizing paraphrasing of smaller units would occur until a fit was achieved, or lack of fit was established.

## Concluding Remarks

The current investigation documented differences in the way examinees solve different kinds of reasoning problems and confirmed that the introduction of additional item types on the GRE analytical measure would broaden the range of reasoning skills assessed. However, given the small sample of items, the small, select sample of examinees, and the exploratory nature of the research, many implications of this study need to corroborated and augmented through further research. In particular, four areas of further research are recommended. The first would concern more detailed analysis of task characteristics and their contribution to correlational structure. Such studies would clarify whether current test development classifications need to be modified. A second area of research would involve an analysis of errors made by examinees who vary in ability. This research would contribute to construct validation and also lay a basis for the use of these kinds of items in instruction. A third area worth investigating is the relationship between disciplinary training and performance on different types of reasoning problems. For example, the close reading of and analysis of claims

characteristic of CV items and some LR items might be influenced by training in the humanities, and generating and evaluating alternative explanations might be influenced by training in the natural and social sciences. Finally, experimental studies of verbal and informal reasoning items will provide an opportunity to develop and test models of reasoning in knowledge-rich but not domain-specific contexts.

## References


Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. Psychological Review, 85, 1-21.

Cattell, R. B., & Burdsal, C. A. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. Multivariate Behavioral Research, 10, 165-179.

Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. Cognitive Science, 13, 1-49.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Dreyfus, H. L. (1992). What computers still can't do. Cambridge, MA: MIT Press.

Emmerich, W., Enright, M. K., Rock, D. A., & Tucker, C. (1991). The development, investigation, and evaluation of new item types for the GRE analytical measure. (GRE Board Professional Rep. No. 87-09P, ETS Research Report No. 91-16). Princeton, NJ: Educational Testing Service.

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. Psychological Bulletin, 105(3), 331-351.

Gellaty, A. R. H. (1989). Human inference. In K. J. Gilhooly (Ed.), Human and machine problem solving. New York: Plenum Press.

Greeno, J. C., & Simon, H. A. (1988). Problem solving and reasoning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), Steven's handbook of experimental psychology: Vol. 2. Learning and cognition. New York: John Wiley & Sons.

Groen, G. J., & Patel, V. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The nature of expertise (pp. 287-310). Hillsdale, NJ: Erlbaum.

Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episodic structure of algebra story problem solving. Cognition and Instruction, 6(3), 223-283.

Hartshorne, C., & Weiss, P. (Eds.). (1931-1958). Collected Papers of Charles Sanders Peirce. Cambridge, MA: Harvard University Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). Induction: Processes of inference, learning, and discovery. Cambridge, MA: MIT Press.

Johnson-Laird, P. (1983). Mental models. Cambridge, MA: Harvard University Press.

Kingston, N. M., & Dorans, N. J. (1982). The effect of the position of an item within a test on item responding behavior: An analysis based on item response theory (GRE Board Professional Rep. No. 79-12bP, ETS Research Report No. 82-22). Princeton, NJ: Educational Testing Service.

Kingston, N. M., & McKinley, R. L. (1988, April). Assessing the structure of the GRE General Test using confirmatory multidimensional item response theory. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kuhn, D. (1991). The skills of argument. New York: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs: NJ: Prentice-Hall.

Powers, D. E., & Swinton, S. S. (1981). Extending the measurement of graduate admission abilities beyond the verbal and quantitative domains. Applied Psychological Measurement, 5(2), 141-158.

Powers, D. E., & Swinton, S. S. (1984). Effects of self-study for coachable test item types. Journal of Educational Psychology, 76, 266-278.

Reimann, P., & Chi, M. (1989). Human expertise. In K. J. Gilhooly (Ed.), Human and machine problem solving. New York: Plenum Press.

Rock, D. A., Bennett, R. E., & Jirele, T. (1988). Factor structure of the Graduate Record Examinations General Test in handicapped and nonhandicapped groups. Journal of Applied Psychology, 73, 383-392.

Schaeffer, G. A., & Kingston, N. M. (1988). Strength of the analytical factor of the GRE General Test in several subgroups: A full-information factor analysis approach (GRE Board Professional Rep. No. 86-7P). Princeton, NJ: Educational Testing Service.

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), Handbook of learning and cognitive processes. Hillsdale, NJ: Erlbaum.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), Eucational measurement. New York: Macmillan.

Swinton, S. S., & Powers, D. E. (1983). A study of the effects of special preparation on GRE analytical scores and item types. Journal of Educational Psychology, 75, 104-115.

Swinton, S. S., Wild, C. L., & Wallmark, M. M. (1982). Investigation of practice effects on types of questions in the Graduate Record Examinations Aptitude Test (GRE Board Professional Rep. No. 80-1cP). Princeton, NJ: Educational Testing Service.

Thagard, P. (1989). Explanatory coherence. Behavioral and Brain Sciences, 12, 435-502.

Toulmin, S. E. (1958). The uses of argument. New York: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). An introduction to reasoning (2nd ed.). New York: Macmillan.

Voss, J. F., Blais, J., Means, M. L., Greene, T. R., & Ahwesh, E. (1989). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. In L. Resnick (Ed.), Knowing, learning and instruction. Hillsdale, NJ: Erlbaum.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research theory (pp. 165-213). New York: Academic Press.

Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.). (1991). Informal reasoning and education. Hillsdale, NJ: Erlbaum.

Ward, W. C., Carlson, S. B., & Woisetschlaeger, E. (1983). Ill-structured problems as multiple-choice items (GRE Board Professional Rep. No. 81-18P). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1985). The relationship of GRE General Test item-type part scores to undergraduate grades (GRE Board Professional Rep. No. 81-22P, ETS Research Report No. 84-38). Princeton, NJ: Educational Testing Service.

## Table 1

## Analytical Reasoning (AR)

Questions 94-98

An airline company is offering a particular group of people two package tours involving eight European cities—London, Madrid, Naples, Oslo, Paris, Rome, Stockholm, and Trieste. While half the group goes on tour 1 to visit five of the cities, the other half will go on tour 2 to visit the other three cities. The group must select the cities to be included in each tour. The selection must conform to the following restrictions:

Madrid cannot be in the same tour as Oslo.

Naples must be in the same tour as Rome.

If tour 1 includes Paris, it must also include London.

If tour 2 includes Stockholm, it cannot include Madrid.

94. Which of the following is an acceptable selection for the two tours?

| | Tour 1 | Tour 2 |
|---|---|---|
| *(A) | Madrid, Naples, Rome Stockholm, Trieste | Paris, London, Oslo |
| (B) | London, Madrid, Paris Rome, Trieste | Naples, Oslo, Stockholm |
| (C) | London, Madrid, Paris | Naples, Oslo, Rome Stockholm, Trieste |

95. If tour 2 includes Rome, which of the following CANNOT be true?

(A) Trieste is in tour 1.

(B) Madrid is in tour 2.

*(C) Stockholm is in tour 2.



96. If tour 2 includes Paris, which of the following must be true?

(A) London is in tour 1.

*(B) Naples is in tour 1.

(C) Stockholm is in tour 2.

97. If tour 1 includes Paris and tour 2 includes Madrid, which of the following must also be included in tour 2?

(A) London

(B) Oslo

*(C) Rome

98. It is <u>impossible</u> for the three cities in which of the following groups to be together in either of the tours?

*(A) Naples, Oslo, and Paris

(B) Oslo, Rome, and Trieste

(C) Paris, Madrid, and Stockholm

## Table 1 (continued)

## Logical Reasoning (LR)

53. A study of the use of the polygraph, or lie detector, found that when a trained examiner using approved questioning techniques gave the test, information from the lie detector was accurate in determining whether responses were truthful for 70 to 90 percent of the responses.

Which of the following conclusions can reliably be drawn on the basis of the results above?

*(A) With a trained examiner using approved questioning techniques, the lie detector failed to give correct results in at least one out of ten instances.

(B) The lie detector gives accurate results only when employed by a trained examiner using approved questioning techniques.

(C) If a trained examiner using approved questioning techniques asks a specific question and the lie detector indicates the answer was false, the respondent definitely lied when giving that answer.

54. If a physician can be confident of a diagnosis of acute illness, the treatment prescribed will be conservative: the minimum expected to aid the patient. This is one reason treatment for a specific illness usually costs more in hospital emergency rooms than in physicians' private offices.

Suppose that the information above is accurate. Each of the following statements, if true, helps to explain why treatment usually costs more in emergency rooms than in physicians' private offices EXCEPT:

(A) Physicians working in their private offices can often rely on knowledge of the patient's history over a period of time.

(B) In emergency rooms, hospital staff unfamiliar with patients who come with severe illnesses often apply several remedies at once to be more certain of obtaining results.

*(C) The variety of illnesses treated by emergency room physicians is much smaller than that treated by physicians in their offices.



6. A man charged with theft of cable television services by making an unauthorized connection said, "They even want restitution of $662 they claim I owe them, which is ridiculous, because I thought some of those shows I saw were awful."

The man's assertion constitutes evidence to show that he

(A) owes the amount the cable service claims he owes

*(B) did watch programs of the cable service

(C) was at some time aware that his hookup to the cable service was unauthorized

7. There is no reason to rule out the possibility of life on Uranus. We must therefore undertake the exploration of that planet.

The argument above assumes that

(A) Uranus is the only other planet in the solar system capable of supporting life

(B) Uranian life would be readily recognizable as life

*(C) the search for life is a sufficient motive for exploration of the planet Uranus



BEST COPY AVAILABLE

## Table 1 (continued)

### Analysis of Explanations (AX)

Questions 86-89

Situation: After serving two terms in the state legislature, Joan Deeker decided to devote more time to writing. However, she knew that it would be difficult to find a job related to politics that would provide both sufficient income and time to write. Since leaving college, she had constantly been involved in politics, first in city elections, and then in her own campaigns. She had introduced a number of liberal social programs and was popular with voters. Since she was likely to win if she ran again, she was also concerned that her decision not to run might hurt her party. When she learned that an appointment in political science at a local university was going to be offered to her and that Louise Jones, a highly qualified candidate, was willing to run in her place, she announced her decision not to run for reelection.

Result: That fall, Deeker ran for her third successive four-year term in the state legislature.

In the context of the situation, the result needs explanation; you will be asked about explanations and statements relevant to explaining the result.

A statement is relevant to explaining the result if there is some possible adequate explanation of the result which the statement either supports or weakens.

Do not consider explanations that are remote and improbable. Borderline judgments about adequacy will not be required.

86. Which of the following statements, if true, is relevant to some possible adequate explanation of the result?

* (A) Prior to the election, Louise Jones suffered serious business reverses.

(B) Deeker's most popular social legislation was directed to the improvement of child-care facilities.

(C) The university appointment in political science became open when a tenured professor suddenly became ill.



87. Which of the following statements, if true, is relevant to some possible adequate explanation of the result?

(A) Deeker's first campaign for a seat in the state legislature was unsuccessful.

(B) The city in which the university is located is a considerable distance from the state capital.

* (C) An organization of teachers sent an investigating committee to look into new charges that the university's policies governing academic freedom were repressive.

88. Which of the following statements, if true, is relevant to some possible adequate explanation of the result?

(A) Deeker was a leading figure in a successful campaign to bring the salaries of the legislators in her state to approximately the same level as the salaries of legislators in nearby states.

(B) The constitution of Deeker's state had once limited to three the number of consecutive terms a state legislator could serve.

* (C) The university learned that Deeker's plans for writing included a book on a highly controversial topic.

89. Which of the following, if true, CANNOT provide the basis for an adequate explanation of the result?

(A) The leaders of Deeker's party convinced her that she could best serve her party by remaining in the legislature and devoting some of the time she had spent on committee work to writing her book.

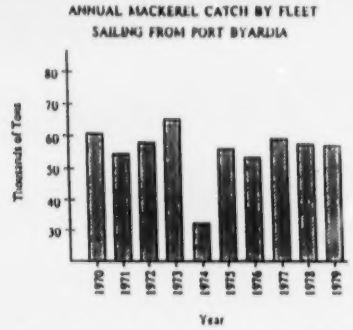
*(B) Deeker and Jones, who had been close friends from the time they first met in college, decided to collaborate on a book about the political history of the state.

(C) Louise Jones took an unpopular stand on a controversial issue, and the leaders of the party convinced Deeker that she was the only one who could win the election.

## Table 1 (continued)

## Numerical Logical Reasoning (NLR)

Questions 37-38 are based on the following graph.




37. Each of the following, if true, provides an adequate explanation for the unusual size of the catch in 1974 EXCEPT:

(A) A major oil spill during the 1974 fishing season temporarily depleted the food supply of the mackerel, leading them to change their feeding grounds for the remainder of the year.

(B) During 1974, fishing fleets of other nations competed for the first time with those from Port Byardia in Port Byardia's traditional fishing areas, but by the next season a treaty had been negotiated reserving those areas for the local Port Byardia fishing fleet.

*(C) Outmoded methods cut down the fleet's effectiveness; after the 1974 season, more modern equipment and methods were introduced.

38. One possible explanation for the aberrant 1974 figure is based on the following:

Between the 1973 fishing season and the 1974 fishing season there was a threefold increase in the price of the diesel oil needed to fuel fishing vessels. The price of mackerel was at the 1973 level for most of 1974, but it rose sharply toward the end of 1974.

Which of the following, if true, best supports an explanation on this basis?

(A) A prolonged strike at the only cannery near Port Byardia eliminated the fishers' outlet for the sale of their catch, so they stopped fishing halfway through the 1974 season.

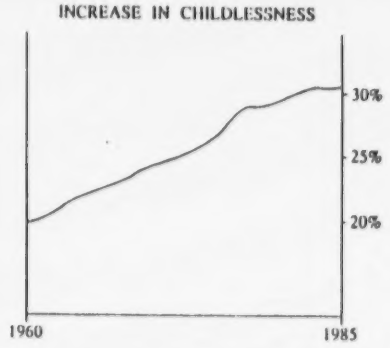
*(B) In 1974, the Port Byardia fleet confined its fishing for mackerel to the areas closest to port.

(C) Unusually severe storms cut drastically the number of days that it was safe for boats to go out fishing during 1974.



Questions 39-40 are based on the following graph.

A study was made of trends in childlessness among women in the United States




39. If the purpose of the study was to determine whether women in the United States have become more likely since 1960 to decide to postpone or forego having children, the data above should be considered along with which of the following?

(A) Annual data for 1960-1985 on the average number of single-parent households

(B) Annual data for 1960-1985 on the average number of children women who are between the ages of 25 and 40 have

*(C) Annual data for 1960-1985 on the percentage of women under 40 who are physically unable to have children

40. Each of the following, if true, could be a factor contributing to the trends indicated in the graph above EXCEPT:

*(A) For each year between 1960 and 1985, a majority of women in the study were over 30 years of age.

(B) Women under 40 were more likely in the 1980's than in the 1960's to devote their time and energy to establishing careers.

(C) The average age at which women first married increased between 1960 and 1985.

**Table 1 (continued)**

Contrasting Views (CV)

Questions 26-29 are based on the following contrasting views.

18th-century view: The new science will liberate the human mind and provide us with a mastery of nature, with which we will break the bonds of tyranny, transform society, and improve all the conditions of life. Rank and birth will fall into contempt in the new age of democratic progress; science is progressive.

20th-century view: Science and technology make possible, not only new products from natural resources, but also new processes of production; not only new techniques of farming, but also new crops. This enables our industry and agriculture to remain competitive. Technical advances will unavoidably result in unemployment and dislocations of the industrial and farm labor force in our society; this is, however, the price of progress.

26. The two views differ most with respect to their conceptions of which of the following?

*(A) Progress

(B) Nature

(C) Society

27. The eighteenth-century view, but not the twentieth-century view, rests on an assumption that

(A) science is value-free and can be used either for good ends or bad

(B) the privileged would invest in technology and would reap the rewards

*(C) human power over nature would be used to benefit people who had held little political power



28. Which of the following, if true, would provide grounds for criticism of the eighteenth-century view but not of the twentieth-century view?

(A) Science provides no basis for any distinction among people that would justify making a distinction according to rank.

*(B) The introduction of new science-based farming methods in some societies has increased the power of the landowning class.

(C) New science-based changes in farming practices have enabled an individual farmer to produce larger crops while working fewer hours than ever before.

29. "Progress is inevitable."

This statement is compatible with (can be true along with)

*(A) each view

(B) the eighteenth-century but not the twentieth-century view

(C) the twentieth-century but not the eighteenth-century view

Table 2a

Exploratory Factor Analysis* of Item Parcels from the GRE General Test and an Experimental Reasoning Test

| Item Types | Item Parcels | Factor Loadings** | | | |
|---|---|---|---|---|---|
| | | Verbal | Informal Reasoning | Formal-Deductive Reasoning | Quant. |
| **General Test (4 & 5 options)** | | | | | |
| Antonyms | a | .94 | -.31 | .08 | -.03 |
| | b | .88 | -.21 | .09 | .06 |
| Analogies | a | .74 | -.04 | .02 | -.01 |
| | b | .72 | -.01 | -.07 | .05 |
| Sentence Completion | a | .71 | .06 | -.07 | .11 |
| | b | .62 | .08 | .01 | .10 |
| Reading Comprehension | a | .54 | .24 | .16 | -.04 |
| | b | .55 | .16 | .32 | -.22 |
| Quantitative Comparisons | a | .07 | .04 | .46 | .38 |
| | b | -.03 | .07 | .44 | .41 |
| Discrete Quantitative | a | .06 | -.06 | .40 | .53 |
| | b | -.14 | .04 | .39 | .53 |
| Data Interpretation | a | .04 | -.05 | -.01 | .82 |
| | b | .02 | -.06 | .03 | .79 |
| Analytical Reasoning | a | .02 | -.04 | .87 | .00 |
| | b | .05 | -.08 | .90 | -.02 |
| | c | .00 | .02 | .79 | .07 |
| Logical Reasoning | a | .27 | .03 | .31 | .08 |
| | b | .38 | .30 | -.10 | .14 |
| | c | .46 | .27 | .06 | -.12 |
| **Experimental Battery (3 options)** | | | | | |
| Analytical Reasoning | a | .06 | .09 | .71 | -.06 |
| | b | .06 | .05 | .70 | .00 |
| Logical Reasoning | a | .29 | .37 | .08 | .02 |
| | b | .27 | .39 | .02 | .22 |
| Numerical Logical Reasoning | a | .11 | .53 | -.09 | .32 |
| | b | .11 | .62 | -.21 | .29 |
| Pattern Identification | a | -.24 | .75 | .33 | .07 |
| | b | -.29 | .76 | .32 | .03 |
| Analysis of Explanation | a | .24 | .63 | -.19 | -.13 |
| | b | .36 | .51 | .09 | -.28 |
| Contrasting Views | a | .57 | .18 | -.15 | .12 |
| | b | .53 | .29 | .02 | -.08 |

*Principal components with Promax Rotation.
**Loadings equal to or greater than .30 are highlighted.

Table 2b

Exploratory Factor Analysis of Item Parcels from the GRE General Test and an Experimental Reasoning Test

| | Inter-Factor Correlations | | | |
|---|---|---|---|---|
| | Verbal | Informal Reasoning | Formal-Deductive Reasoning | Quant. |
| Verbal | 1.00 | .59 | .51 | .47 |
| Informal Reasoning | .59 | 1.00 | .60 | .51 |
| Formal-Deductive Reasoning | .51 | .60 | 1.00 | .62 |
| Quantitative | .47 | .51 | .62 | 1.00 |
| Factor Variance Explained | 6.51 | 3.39 | 5.45 | 3.12 |

Table 3

Examples of the Kinds of Elaborative Comments Included in Examinees' Summaries of Problem Situations

Informal inferences

AX: However, the result is that Deeker actually must have changed her mind.

NLR38: What this is saying is that a significant reason why less was caught in '74 was that less people were out catching then because the price of oil had gone up and people were not necessarily willing to spend money for diesel or the price of mackerel had not increased by the end of '74 when the price of mackerel...the ratio sort of got to what it was before the diesel oil increased and was profitable again to spend extra money for oil and still catch mackerel and sell it at the higher price.

Interpretive generalizations

CV: And so, in this, in the first set, progress is positive. And in the second set it's also positive and negative, so the price of progress is negative.

LR53: Okay, so there's accuracy with a trained professional.

Possible Explanations

AX...so it could be, thinking ahead, that maybe she couldn't find a job, and she had to, the job offer fell through, and she had to run again because she needed the money.

AX...so her ego may have kicked in, apparently, so she decide well, she wants to go back just because she doesn't want anyone else taking her place...

Background Knowledge

CV: Sounds like an Enlightenment project....sounds like more of a modified Enlightenment project.

CV: And what's going to happen is that there be ways of farming new crops; industry and agriculture remain competitive; there's a large economic side, capitalistic side to this.

## Table 4

### Descriptions of Justification Categories and Examples of Examinees' Justifications (italics) and Evaluations (underlined)

List - selected phrase or phrases, a proposition or series of propositions are cited or listed to support ar evaluation but the relationships or connections among phrases or propositions are not articulated.

AX87(A) Deeker's first campaign for a seat in the state legislature was unsuccessful.

*No, she knew she'd get elected, she will, but she didn't want to hurt her party, wanted to make sure someone would fill her shoes, wanted a job that could pay,* first campaign is irrelevant

CV28(A) Science provides no basis for any distinction among people that would justify making a distinction according to rank.

*...The 18th century view claims that it will break bonds of tyranny, and that rank and birth will fall into contempt in the new age of democratic progress,* so this is not right.

Interpretation - propositions are related through similarity in meaning or shared concepts or paraphrased to illustrate similarities or differences in meaning.

CV27(B) The privileged would invest in technology and would reap the rewards.

*No, the 18th century view doesn't talk about the privileged. The privileged for them would have been the rank...*

LR6(B) did watch programs of the cable service

which is asserted *because he claims that some of those shows that he saw--which means that he did actually watch the programs--were awful..*

Generalization - statements that summarize information in more general terms or reflect the gist of a passage.

CV26(A) progress

*...Well, this one pretty much views progress as something which is good in itself but only has benefits. This one here views progress as something you need but it also has drawbacks as well....*I'm going to guess it's progress now but I'm going to come back to it later...

LR53(B) The lie detector gives accurate results only when employed by a trained examiner using approved questioning techniques.

(B) seems to be what the passage is trying to state. A trained examiner using approved questioning techniques gave the test, *the lie detector seemed to work pretty well.*

Temporal agreement - temporal conjunctions or disjunctions between events are noted.

AX88(B) The constitution of Deeker's state had once limited to three the number of consecutive terms a legislator could serve.

*But that was, it seems like, well this does apply to the past...*it doesn't seem relevant to the present situation.

AX86(C) The university appointment in political science became open when a tenured professor suddenly became ill.
again. (C) doesn't really explain Joan's decision to go back and campaign. *This happened before Joan decided not to run again.*

Informal inferences - inferences based on background knowledge

LR54(A) Physicians working in their private offices can often rely on knowledge of the patient's history over time.

*Therefore they can be more confident of an illness.*

AX87(C) An organization of teachers sent an investigating committee to look into new charges that the university' policies governing academic freedom were repressive.

But (C)--*If she looked upon the university frowningly, if she rejected the offer, then it would make sense of the result that she ran again...*and that would be (C).

BEST COPY AVAILABLE

Table 4 (continued)

Descriptions of Justification Categories and Examples of Examinees' Justifications (italics) and Evaluations (underlined)

Suppositions - speculative scenarios about possible antecedents and consequences. Often similar to informal inferences except that they have a more tentative or hypothetical flavor keyed by words such as "might," "maybe," "perhaps," "unless."

AX87(B) City in which the university is located is a considerable distance from the state capital.

*Unless she wanted to keep her hand in, then perhaps she wouldn't want to take that job and decided to run again.*

AX86(A) Prior to the election, Louise Jones suffered serious business reverses.

I would say (A) *because if she suffered serious business reverses, it might have hurt the political career.*

Rule-based - consistency with a specific, definite rule is the basis for option evaluation.

AR94(B) London, Madrid, Paris, Rome, Trieste.

*Well Paris has London, that's fine. Madrid, that's okay. Naples - not Rome,* no good.

AR94(C) London, Madrid, Paris.

*If tour 1 would be London, Madrid, and Paris, that only has 3 cities and tour 1 has to visit 5 of the cities.* So that's not the answer either.

Formal Deduction 1 - step-by-step inference based on specific, definite rules that follow from a given proposition.

AR98(A) Naples, Oslo, and Paris.

*Naples needs Rome in the first one. Paris needs London. So that gives us Stockholm, Trieste, and Madrid.* That's the one.



Formal Deduction 2 - step-by-step inference based on specific, definite rules that follow from the contrary of a given proposition.

AR96(C) Stockholm is in tour 2.

Let's see if we can adjust this. *Stockholm is in tour 1, it can be with anything. Just that Oslo can't be with Madrid so we'll leave Oslo there. We'll leave Madrid here. London, Naples, Rome and Trieste can be there.* So that doesn't have to be true.

Match - option is justified by comparing it to the outcome of a reasoning episode that occurred prior to reading the options.

[prior summary of stimulus- *So opposed to the 18th century view the 20th century person doesn't see progress as leading to a better society...but it improves production...and so it's worth the cost.*

CV.26 The two views differ most with respect to their conception of which of the following?

*Progress, because the 20th century sees progress in production and social something or other as being probably divergent, and the 18th-century person sees it going hand-in-hand.* So that's (C).

AR97. If tour 1 includes Paris and tour 2 includes Madrid...

*1 includes Paris then it also has London, and tour 2 includes Madrid that means Oslo is here....then Rome has to be here and Naples. Because Oslo and Madrid have to be separated...and London has to be in 1 with Paris.* So that's Rome.

None - Either no statement of a reason for selecting or rejecting an option is given or the stated reason is a simple restatement of the option and the stem.

CV28(C) New science-based changes in farming practices have enabled an individual farmer to produce larger crops while working fewer hours than ever before.

This is not, this is agrees with the 18th century view and it does not provide grounds for criticism.

Unclassifiable - remarks that were unintelligible, uninterpretable, or did not fit any of the above categories.

Table 5

Proportional of Responses Classified into Different Evaluation/Justification Categories for Various Item Types

| Item Type | No Explicit Evaluation or Justification | Evaluation Only | Justification Only | Evaluation Followed by Justification | Justification Followed by Evaluation |
|---|---|---|---|---|---|
| CV | .30 | .19 | .06 | .31 | .14 |
| LR | .18 | .32 | .04 | .35 | .10 |
| NLR | .11 | .20 | .14 | .42 | .13 |
| AX | .09 | .34 | .10 | .34 | .13 |
| AR | .17 | .03 | .10 | .10 | .60 |

Table 6

Proportional Frequency of Various Kinds of Justifications for each Item

| | | | Justification Categories* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | P+** | N/U | L | I | G | T | II | S | FD1 | FD2 | M | R |
| CV26 | 60.0 | .35 | .37 | .03 | .54 | | | | | | .06 | |
| CV27 | 86.2 | .32 | .47 | .18 | .26 | | .09 | | | | | |
| CV28 | 60.3 | .39 | .21 | .39 | .30 | | .09 | | | | | |
| CV29 | 72.7 | .67 | .18 | .12 | .59 | | .12 | | | | | |
| LR 6 | 87.8 | .71 | .21 | .50 | | | .29 | | | | | |
| LR 7 | 94.1 | .54 | .14 | .82 | | | .05 | | | | | |
| LR53 | 84.6 | .32 | .29 | .50 | .15 | | .06 | | | | | |
| LR54 | 68.4 | .41 | .10 | .10 | .03 | | .73 | .03 | | | | |
| NL37 | 72.4 | .39 | .03 | .07 | | .03 | .63 | .23 | | | | |
| NL38 | 76.2 | .15 | .43 | .02 | .02 | | .39 | .14 | | | | |
| NL39 | 66.2 | .43 | .07 | .25 | .04 | | .61 | .04 | | | | |
| NL40 | 63.0 | .35 | .03 | .18 | | | .58 | .21 | | | | |
| AX86 | 87.3 | .35 | .12 | .18 | | .06 | .26 | .38 | | | | |
| AX87 | 68.4 | .27 | .24 | .03 | | .05 | .24 | .45 | | | | |
| AX88 | 85.4 | .33 | .12 | .06 | .03 | .06 | .24 | .41 | .03 | | | .06 |
| AX89 | 85.1 | .77 | .18 | .09 | | | .64 | .09 | | | | |
| AR94 | 64.3 | | | | | | | | | | | 1.00 |
| AR95 | 51.4 | .14 | .02 | | | | | | .48 | .02 | .38 | .10 |
| AR96 | 58.1 | .19 | | | | | | | .48 | .24 | .21 | .07 |
| AR97 | 85.4 | .35 | | | | | | | .10 | | .81 | .10 |
| AR98 | 42.7 | .17 | .02 | | | | | | .78 | | .04 | .16 |

**Key:**

| | | | | | |
|---|---|---|---|---|---|
| N/U - | None or Unclassified | T - | Temporal Agreement | FD 2 - | Formal Deduction 2 |
| L - | List | II - | Informal Inferences | M - | Match |
| I - | Interpretation | S - | Suppositions | R - | Rule based |
| G - | Generalization | FD1 - | Formal Deduction 1 | | |

* Two most common categories highlighted for each item type.

** Percent correct from Emmerich et al. (1991).

Figure 1

One Examinee's Diagram for a Practice Contrasting Views Problem Set

Questions 69-73 are based on the following contrasting views.

View I: A painting's form--its use of line, color, and shape--arouses the viewer's aesthetic sense, whereas its content, if appealing or interesting, often interferes with the viewer's aesthetic appreciation. Abstract masterpieces lacking discernible subjects, because they provide a source of pure aesthetic experience, as opposed to sentimental or intellectual experience, are the highest form of art.

View II: Art engages the mind, inspires the soul, and arouses the senses. In great art, form and content cooperate perfectly, so that the eye, stimulated by the formal beauties of line, color, and shape, lingers to search out the deeper truth of what it sees. Aesthetic experience satisfies so deeply precisely because it involves all our faculties, sensory, intellectual, and spiritual.

BEST COPY AVAILABLE

54020-12395 • Y35M 550 • 209084

BEST COPY AVAILABLE